**Sunil Sah**

**Project 3b Retail Biz Segmentation Report**

**Clustering Analysis of the Mall Customers Dataset: Report**

This analysis demonstrates a step-by-step application of clustering techniques to segment customers in the "Mall Customers" dataset. The dataset was loaded from an online source and explored to understand its structure and statistical properties. The initial inspection revealed no missing values, ensuring data integrity for further analysis.

Key features selected for clustering include Age, Annual Income (k$), and Spending Score (1-100). These attributes were standardized using StandardScaler to ensure that differences in scale did not disproportionately influence the clustering process. Exploratory visualizations showed the distribution of these features, highlighting distinct patterns in age groups, income ranges, and spending behavior.

To determine the optimal number of clusters, we applied the Elbow Method and calculated Silhouette Scores. The Elbow Method, which evaluates the within-cluster sum of squares (inertia), suggested diminishing returns beyond five clusters. Silhouette Scores validated this choice by measuring the separation distance between clusters, with a peak score observed at five clusters.

The K-means clustering algorithm was applied with five clusters, and each data point was assigned to a respective cluster. The results were visualized using a scatter plot that mapped Annual Income against Spending Score, with each cluster represented by a unique color. The visualization revealed distinct customer groups characterized by varying income and spending patterns.

Cluster-level insights were derived by examining descriptive statistics for each group. For instance:

* Cluster 0: Customers with low income and low spending scores, representing a cautious spender group.
* Cluster 1: High-income, high-spending customers, likely luxury shoppers.
* Cluster 2: Moderate-income customers with high spending scores, indicating potential impulsive buyers.
* Cluster 3: High-income customers with low spending scores, suggesting a savings-oriented group.
* Cluster 4: Low-income customers with moderate spending, showing balanced behavior.

This segmentation enables tailored marketing strategies, such as offering loyalty rewards to high spenders, targeted promotions for moderate spenders, or budgeting tools for cautious spenders.

Future improvements could involve incorporating advanced clustering techniques like hierarchical clustering or DBSCAN for non-linear patterns, integrating additional data (e.g., transaction history), and automating the analysis pipeline for dynamic updates. These enhancements would provide deeper insights and more personalized customer engagement strategies.